# Morphometric elements of urban form

*The content of this chapter was partially published in @fleischmann2020.*

The first half of the thesis presented the work done in the field of urban morphometrics to date, used as a basis for the formulation of hypothesis and research questions made in the previous chapter. The following chapter 6 is the first of the three chapters which focus on operationalising top-level ideas and answering questions posed in chapter 5.

The aim of this chapter is to lay the building blocks for morphometric analysis out. It focuses on morphometric elements of urban form, their relation to traditional morphological elements and range of possible issues from conceptual to practical ones. To overcome some of them, this chapter introduces the concept of morphological tessellation as the smallest spatial unit of analysis. The critical question which needs to be answered is the way of combining all elements together, as literature offers a range of potential frameworks, but few of them are explicitly defined, and minimum has the ability to link all elements and their aggregations together.

Structurally, the following chapter is divided into five major sections. Section starts with the introduction of analytical representation of urban form, covering fundamental morphometric elements of urban form and their aggregations into larger features. The specific focus is on the issues related to the plot as the smallest spatial unit and potential use of morphological tessellation. In its second part, the section follows with the overview of analytical frameworks (), i.e. the ways how to combine elements and their aggregations into a singular schema. Section builds on the previous in a proposal of a relational framework of urban form for comprehensive morphometric applications, relying on topology and morphological tessellation. To ensure that tessellation holds as a spatial unit, the rest of the chapter is dedicated to empirical experiments assessing its ability to capture the similar information as the plot and the potential of topological aggregation in comparison with other location-based methods. Hence section outlines the method of testing and section presents the results of the empirical study. The chapter concludes with a short discussion of presented results and paves the way to the application of relational framework and tessellation in chapter 7.

## Analytical representation of urban form

This chapter aims to talk about the analytical, measurable representation of urban form. It does not attempt to build a new theory of urban morphology or re-conceptualise its understanding. It focuses on the operationalisation of the analysis within the context of urban morphometrics. For that, we need to understand two aspects - what are the basic components of urban form we can capture in relevant data, and how they relate to each other. In other words, what are the morphometric elements of urban form and how to work with them in an analytical framework.

### Elements of urban form in the context of urban morphometrics

Morphological theory talks about a wide variety of elements, but only some are seen as fundamental. Furthermore, we have to take into account the issue of data availability - since high-quality data representing urban form are not always available, we need to use as a small number of resources as possible to keep any morphometric method widely applicable. For this reason, this research attempts to work with the minimal input needed - fundamental elements only.

#### Fundamental morphometric elements

Urban morphology knows three fundamental elements of urban form. Both [@moudon1997] and later [@kropf2017] agree on buildings, streets and plots as features representing three fundamental aspects of human behaviour: *sleep*, happening in a shelter, i.e. building, *movement*, first reflected in tracks and today in streets and roads and *local activity*, initially happening in a core territory, which can be today seen as a plot [@kropf2017].

Fundamental *morphometric* elements are based on fundamental morphological elements, and their goal is to reflect the same phenomena. However, the difference between morphological and morphometric elements is that the latter are operational, measurable and represented as (usually vector) GIS data. Hence we can talk about three layers of data: buildings (either as footprints or models), street networks and plots as smallest spatial units.

It is essential to ensure that the data are good enough to represent morphometric elements. That could be an issue for all types of elements, so there are cases when the data needs to be prepared for morphometric analysis. The pre-processing can be in some cases automatised, in other, unfortunately, manual or at least semi-manual to have the data in a correct form in the end.

Whilst each dataset coming from a different source is specific; hence the cleaning procedure needs to be tailored to each source, there are some common issues which are not unique to specific datasets. The following section presents all fundamental morphometric elements and outlines these common issues and ways of resolving them or at least minimising the error under a significant level. We cannot expect data to be perfect all the time (they are never perfect). This limitation has to be taken into account during the design of any morphometric analysis, which needs a certain level of robustness to accommodate potentially erroneous data.

##### Buildings

Buildings can be represented as footprints, i.e. two-dimensional projections of building shape to the ground or 3D models. Somewhere in between lies what we can call the 2.5D model, which is a building footprint with an attribute of a building height. All the options are being used in morphometric research. However, since one of the key goals of the method being developed within this work is wide applicability, it cannot depend on 3D representation due to its limited availability. The most accessible representation of buildings are vector-based footprints (figure ), either generated by official national or municipal mapping agencies (like Ordnance Survey in Great Britain) or by volunteers within OpenStreetMap crowdsourcing mapping movement [@haklay2008openstreetmap]. Hence this research depends on 2.5D vector representation, providing a balance between availability and precision.

Building footprints in the optimal resolution and data quality.

Having data layer representing building footprints correctly and consistently is the first condition for successful morphometric analysis. There are several aspects which need to be fulfilled - topological correctness, consistency in detail, representation of individual buildings and building height attribute presence. Overall, it is expected to have a building data representing Level of Detail 1 (LoD1) [@biljecki2016] (figure ).

Diagram of LoD classification on the example of single-family housing. (Biljecki et al. 2016, figure 3)

Topological correctness ensures that geometry represents the relationship between buildings on the ground. There are characters measuring continuity of a perceived wall in a joined buildings or shared walls ratio which require building polygons to be correctly snapped together when two buildings touch. In that case, it is expected that neighbouring polygons will share vertices and boundary segments. There should not be a gap between polygons when there is none in reality. Also, polygons must not overlap at any case as that could cause significant disruption of analysis.

The building detail should be consistent across the dataset and represent an optimal approximation of building shape based on LOD specification as proposed by @biljecki2016. The approximation should represent LOD1.1 (no details, but the shape is kept) or LOD 1.2 (minor details), building shapes should not be overly detailed nor overly simplified. In the case of inconsistency, simplification of more detailed shapes should be done before morphometric assessment.

Each polygon should represent a single building. There are datasets (often of a remote sensing origin) capturing all structures which are joined by any means as a single polygon. Such data do not represent the morphological truth on the ground. Their pre-processing is complicated as it requires splitting of existing geometries based on an additional dataset. The second extreme is the opposite situation when multiple polygons represent a single building. These usually represent different height levels, through routes or similar features. If these polygons, representing parts of buildings, have a common ID which allows joining them together to get a single polygon representing a single building, the pre-processing of such a data is only a simple dissolution. However, there are many cases when this ID is missing, and correct pre-processing require either clever heuristics to understand which polygon belongs to which or laborious manual work.

A number of morphometric characters use building height attribute, which, in that case, has to be present in the original input dataset. The resolution should be able to capture the distinction between floor levels.

##### Street network

Street network represents a street, but that itself can be, form the data perspective, defined as multiple features. We can understand it as a movement, as an area or as a network. Furthermore, an important aspect of a street (network) is also a junction - node relationship. Therefore, within this study, a street is represented as a street network consisting of edges representing centrelines and nodes representing junctions. Both can be abstracted to a simple LineString (polyline) geometry and its configurational graph representation.

Illustration of street network represented as street centrelines, from which can be derived both street edge and node.

The similar pre-processing situation as with building layer is with a street network. Incorrectly drawn street network may cause significant errors in morphometric results and consequently in taxonomy. There are three critical cases which need to be checked before the analysis - topological correctness, morphological correctness and consistency in classification.

Topological correctness ensures that each street segments is represented by a single LineString geometry, that neighbouring segments share end vertex and that geometry is not split if the segments intersect only on the projected plane and not in reality (typically multi-level communications, when one is on the bridge across the other so that projected intersection is not a real intersection).

Moreover, street networks have to be morphologically correct, which means that geometries represent morphological connections, not other, usually transport-focused elements. That often mean simplifications of networks to eliminate transport geometries like roundabouts or similar types of junctions, or dual lines representing dual carriageways. In certain cases, networks have to be snapped together, because due to traffic-calming measures, some junctions might not be connected when they should be (morphologically).

Finally, the network needs to be consistently drawn in terms of inclusion of different levels of network hierarchy. Hence the definition of what is considered a street and what is a minor pedestrian connection is crucial and needs to be consistent throughout the study.

Subject to data availability, networks are widely available. However, geometries mostly represent transport network and often do not follow ideal topological rules. The pre-processing to ensure that all three points above are fulfilled is hence necessary and can be partially automatised either using momepy or using methodology outlined by @krenz2018, using conventional GIS tools. However, there might be cases when more complicated procedures should be employed, either to provide a more accurate algorithm or to include manual steps.

##### Spatial unit

The identification of a reliable, significant and universal spatial unit of analysis is of crucial importance. However, the situation with a traditionally used plot is complicated.

###### Plot

In traditional Urban Morphology, a *plot*, is considered to be the smallest meaningful unit of spatial subdivision and a fundamental component to understanding the spatial structure of the ordinary fabric of urban settlements [@panerai2004; @porta2014; @moudon1997] and their processes of formation and transformation in time [@whitehand1981].

However, despite its significance, the plot remains a problematic construct. At *ontological* level, there is no agreement on *what exactly a plot is*: indeed, it has been variously defined as “*a land-use unit defined by boundaries on the ground*” [@conzen1969, p. 128], *a module* of the urban tissue constituted by a built-up area and its open pertinent area [@caniggia1979], a piece of property, subject to subdivision and amalgamation as a result of successive patterns of occupation [@moudon1986], or again, according to @bobkova2017, as “*a basic unit of control*”, “*a fundamental link between spatial and non-spatial medium*”, “*a connection between built space and space of movement*” and “*the framework for building evolution over time*” (p. 47.5). Furthermore, crucially, more often than not, these definitions may represent very different entities on the ground “*potentially leading to misinterpretation and so a somewhat obscured picture of the dynamics of urban form*” [@kropf1997, p.1].

The second problem has to do with the relevance and applicability of the plot to different urban contexts. In literature, plots have been predominantly used to study and characterise traditional urban tissues that having evolved incrementally at the plot level [@bobkova2017; @conzen1969], are quintessentially plot-based [@panerai2004]. It is, however, not the case for urban forms that came about after the Second World War, which appear to respond to substantially different rules of the organisation [@dibble2016; @feliciotti2018]. For these tissues, “*plots no longer have a structuring role*” [@levy1999 p.83], and hence can hardly be a suitable unit of analysis. While the process of identifying plots in traditional tissues is somewhat less controversial, the same is not true in contemporary ones (figure ).

Comparison of traditional (left) and modernist (right) urban tissues in Glasgow. Plots are clearly better identifiable—even just visually—in the former, where distinction of public and private space is clear-cut, than in a modernist housing estate, where the transition between public and private is blurred. Source: Ordnance Survey MasterMap, January 2019 (EDINA Digimap Service)

In addition to this issue, the identification of plots in the urban fabric also poses a series of *analytical* problems: given a map or a satellite image, how to determine the plot boundaries consistently? Moreover, in the case of existing datasets, what do they *actually* represent? What definition of plots do they adopt? Are different datasets comparable?

Not all mapping agencies explicitly report plots and, even when they do, not all of them define or represent plots in the same way. In some spatial databases, as in the Swiss Katasterwesens, plots are represented as a unitary land parcel, whilst in other cases, ownership-based plots can be made of multiple unlinked features, as in the French Cadastre, limiting comparability between different datasets. In other cases, the identification of plots from available sources is inferred by the analyst via resource-intensive manual interpretation. However, that makes the resulting procedure on the one hand unsuitable for large scale analysis, and on the other potentially biased, as heavily dependent on both individual interpretation and the often uneven quality of the underlying data. Indeed, while through open-data policies [@huijboom2011] and Voluntary Geographical Information System (VGIS) [@barrington-leigh2017] the availability of free-to-use geo-data is growing dramatically, their quality, coverage and resolution are often insufficient to determine individual plots and generally limited to building footprints, street centrelines, natural features and administrative boundaries. All of this reduces the reliability of the analysis and the universality of its results considerably.

Given the aforementioned issues, and despite plots being still widely used in urban morphology to capture the “*pattern of human intention and activity*” [@kropf1997, p.5], they are ill-suited as a basic unit for morphometric applications.

###### Morphological tessellation

One of the few alternatives of plots proposed in the literature [@hamaina2012a; @hamaina2013; @schirmer2015; @schirmer2019] is morphological tessellation (MT). A method of deriving a spatial unit of analysis, the morphological cell (MC), which is able to convey reliable, universal and meaningful plot-scale information and, at the same time, to minimise manual labour, subjective interpretation and data dependence. Hence it is proposed to use morphological tessellation as a spatial unit instead of the plot in urban morphometrics.

Illustration of morphological tessellation as the smallest spatial unit with the ability to partially replace the plot while being fully derived from building footprint data.

At the core of the proposed implementation of MT lies the *Voronoi tessellation* (VT), a method of geometric partitioning that from a planar set of *“seeds”* generates a series of polygons, known as *Voronoi Cells* (VC). Each Voronoi cell encloses the portion of the plane that is closer to its seed than to any other (figure ), representing its “*influence zone*”.

Voronoi tessellation based on randomised seeds. Each colour represents the area of one tessellation cell (influence zone). Dashed lines end in infinity

Voronoi tessellation has been already used in relation to urban form, in the context of spatial clustering algorithms [@dogrusoz2007] and built-form geometry generalisation techniques [@ai2007; @basaraner2004; @li2004; @liu2014], or as input for image-based pattern recognition [@yu2017]. In recent years, morphological tessellation was used to study the micro-scale properties of the urban fabric [@hamaina2012a; @hamaina2013] in order to produce a reliable method for urban form patterns’ recognition, which pioneered the generation of VC from building footprints. Later, Schirmer and Axhausen [-@schirmer2015; -@schirmer2019] devised a method to define “*influence zones*” around buildings using a “*topological skeleton*” of unbuilt space that is mathematically similar to morphological tessellation. In parallel, Usui and Asami [-@usui2013; -@usui2017; -@usui2019] included the street network as an additional input alongside the building footprint to the Voronoi tessellation algorithm, to mimic the plot structure of traditional Japanese urban fabrics. Whilst the generated mesh shows remarkable similarity to the plot pattern, its main limitation is the inability to capture the spatial pattern of modernist (post-WWII) urban tissues and the highly variable distance between building and street that is typical of such fabrics. On a similar vein, Araldi and Fusco [-@araldi2017; -@araldi2019] developed an approach based on Voronoi tessellation and street segments to define a spatial unit based on the pedestrian point of view.

In all these cases, the use of Voronoi tessellation helped to rigorously and reliably cluster components according to their configuration although, as pointed out by @usui2019, the relationship between morphological cell and “conventional” plots has never been directly tested to date. In this sense, the morphological tessellation approach is to be intended as a continuation of this line of works, and insofar it too utilises the Voronoi tessellation procedure. However, unlike previous studies, this research aims to provide a fully operational and replicable method by examining the details of the tessellation process and its parameters and testing the similarity of morphometric characters as measured on both morphological cells and plots through direct comparison.

#### Analytical aggregations

Since morphological analysis aims to capture patterns of urban form, it must describe single elements as well as their spatial configurations and relationships. Therefore, larger analytical units have to be identified. Generally, we can distinguish two approaches of aggregation of fundamental elements into larger units: area-based and location-based [@berghauserpont2014].

##### Area-based

Area-based approaches divide space into preselected units, i.e. administrative boundaries [@gielen2017], abstract projected boundaries (grid) [@galster2001], or larger morphological structures such as a block [@gil2012] or a Sanctuary Area [@dibble2017]. However, such methods may face two connected issues, together named “Modifiable Areal Unit Problem” (MAUP) [@openshaw1984]: scale issue (how big the area of aggregation should be) and aggregation issue (where should we draw its boundaries). Specific non-morphological area-based approaches are prone to both of them, particularly the latter: a change of the boundary, for example, the voting district, might affect the analysis’ results.

However, there is still a significant scope to study morphological aggregations. Starting from the smallest plot scale, we recognise street edge, block, street, up to either urban tissue if we follow one definition [@kropf2017] or in sanctuary area if we follow another [@feliciotti2016; @mehaffy2010].

The smallest morphological aggregation of plots is a street edge, which is a “*series of one or more plots served by the same street*” [@feliciotti2016, p.5]. To resolve the issue of corner plots, @feliciotti2016 add “*bound to the centrality of the street is sits on*”, which can be translated as the importance of each street. A corner building simply belongs to the more important street of the two. More massive related aggregation is a block, or what @kropf2017 calls a plot series. Block can be defined “*an aggregate of plots surrounded on all sides by street spaces*” [@kropf2017, p.47]. Both street edges and blocks are combinations of plots, limited by a street network and as such, both are relatively easy to define even algorithmically given appropriate data.

On the larger scale, area-based aggregations result either in Sanctuary area [@dibble2017; @feliciotti2016; @mehaffy2010] or urban tissue, morphological region or another homogeneity-based structure (e.g. urban structural unit, character area). Sub-optimality of Sanctuary area has been explained in chapter 3 and hence is not scrutinised again here.

Urban tissue (figure ) and related concepts, on the other hand, are worth considering as they repeatedly emerge from various schools of urban morphology. As a *morphological region*, or *plan unit*, it has a prominent role in Conzenian historic-geographical approach [@oliveira2020]. While different terms often capture different concepts, the underlying logic is always the same. It is well summarised in Kropf’s definition of urban tissue: “*a distinct area of a settlement in all three dimensions, characterised by a unique combination of streets, blocks/plot series, plots, buildings, structures and materials and usually the result of a distinct process of formation at a particular time or period.*” [@kropf2017, p.89]

Urban tissues in the fringe (top) and central areas (bottom) of Bath identified by Kropf 2017 (figure 8.4)

While some of the area-based aggregations, mostly those following morphological nature of form (edge, block, tissue), are entirely meaningful and minimise MAUP, others are not ideal and should be generally avoided (administrative units, census blocks, voting districts).

##### Location-based

Location-based approaches generate analytical units independently for each source-element as a unique aggregation around it, typically at walking or driving distance, where distance is measured either along with the street network (network distance) or an approximation of it (for example as the crow flies). Therefore, the aggregated values are uniquely and consistently generated for each source-element (e.g. building), and the effect of arbitrary data aggregation is minimised, resolving MAUP’s aggregation issue. For this reason, literature prefers location-based analytical units, as their nature partially resolves MAUP. The scale problem part of MAUP is present also in location-based methods, and it is up to the methodology adopted on a case-by-case base to limit its effect to a minimum.

Currently, morphological literature relies on a few methods to define an aggregation of units in a location-based manner. The most straightforward is based on simple Euclidean (as-crow-flies) distance from the elements of analysis (typically a radius of 400 metres around a building) [@schirmer2015]. However, such an approach does not reflect the actual morphological situation on the ground. In some instances, as in traditional compact urban tissues, it can capture hundreds of buildings within 400 metres. However, only a few in sparse modernist urban tissues, leading to fundamental differences in the amount of information captured, causing issues of comparability of such information.

Excluding the effect of specific tissue types from the definition of aggregations overcomes method based on reach. Following the street network or axial map of urban form [@berghauserpont2014; @marcus2017], it captures the area which is possible to reach within a set distance (mostly metric). As a location-based method, reach is useful because it reacts to unequal morphologies, but only through constraints that limit access to space, rather than through detection of a difference in urban form itself. The logic is based on the cognitive experience of cities but limited to accessible open spaces, excluding the intra-block relationships. It can generate situations of two buildings facing each other across the block (hence directly influencing each other) not being aggregated together, ignoring their relationship. Furthermore, unlike any other method, network-based reach adds a requirement of street network data input limiting its applicability, e.g. in the context of remote sensing-based building footprints in the informal context where no street network is available.

Both Euclidean distance and metric reach methods cannot capture the change in the granularity of urban tissues, hence are effectively measuring different information in granular and sparse tissues. In the case of reach, the distance could be defined topologically as a number of steps on the network (represented by a graph structure) [@berghauserpont2014], allowing to recognise the change in the pattern of aggregations, but it still does not eliminate the issue of intra-block relationships. On top of that, network-based methods face issues in data availability - street networks usually need significant adaptations before they can be used, as they are typically drawn for traffic purposes, not morphological ones; axial maps are scarcely available, and their generation needs a very specialised type of morphological knowledge limiting the applicability of such method.

The third method present in literature is K-nearest neighbour (KNN) analysis, which is based on Euclidean distance, but it is using it differently. It defines an aggregation as a set number of nearest neighbours, defined via as-crow-flies distance. Whilst only scarcely used in urban morphology [@liqiang2013], it has potential as such an approach might reflect changes in the granularity of urban tissues. However, due to the Euclidean definition of nearest neighbours, it cannot react to the detail of some spatial configurations (e.g., be able to detect linear patterns with natural boundaries between as features across the boundary might be closer than those within the pattern). Theoretically, KNN could be used together with reach analysis, joining both the ability to capture morphology represented by networks and scalability of KNN, but this concept has not been applied in morphological literature so far. However, it would still not resolve the issue of intra-block relationship.

It should be pointed out that the morphological cells do offer added values that are relevant on their own in the context of aggregation, regardless of their similarity to the plots. These have to do with the potential innovations – yet largely unexplored – which are triggered by the very nature of this geometry. For example, unlike other methods of urban form partitioning, the morphological tessellation covers the totality of space uniformly within the set study area, allowing to capture the topology of contiguous space at the plot-level. Indeed, since all MCs are determined by adjacency, by using morphological tessellation it is possible to think in terms of topological distance (set number of topological steps between cells) rather than geographic distance (set metric distance around elements, either “as the crow flies” or along with the street network). Moreover, thinking in terms of topological distance as opposed to metric, the morphological tessellation can be used to define new aggregated analytical units that are able to capture the immediate area of influence of a building on its surrounding fabric and, at the same time, of the surrounding fabric on the building. Indeed, since the size of each morphological cells depends on the granularity of the urban structure, the spatial representation of a set topological distance would be far smaller for a morphological cell located in a fine-grained built-up area than for the same located in a coarse one (figure ). Crucially, this is a kind of information that would not be possible to access with plots alone, which allows for reframing the very idea of “proximity” by rethinking the relationship between scale and spatial meaning, thereby enhancing the ability to capture the context in morphometric analysis.

Relationship between morphological cells of topological distance 2: the red geometry represents the adjacency network of neighbouring elements (buildings, morphological cellss) at topological distance 1 (adjacent neighbour of first order), while blue geometry represents the boundary of the aggregated analytical unit of topological distance 2 for each of the highlighted buildings. In the image, a fabric characterised by fewer and sparser buildings (b) generate larger cells and aggregated units compared to a denser and more compact fabric (a).

Topology captures the information on adjacency of neighbouring elements (cells) - two cells are neighbouring if they share at least one point (so-called Queen contiguity) or one segment (so-called Rook contiguity). It defines the proximity of elements in terms of the number of steps needed to get from each element A to each element B. Topological relationships can be of two types - unconstrained, if not limited by any other element than tessellation itself, and constrained if the step between two neighbours is impeded by constraint (a block is the maximum number of topological steps from element without the need to cross the street network, while the street network is the constraint in this case). Thus, we can define an aggregation around each element based on several topological steps (topological reach) on the morphological tessellation, where aggregation defined by n steps includes all morphological cells which we can reach within x <= n steps.

#### Operational Taxonomic Unit as a morphological aggregation

Having defined aggregation models, the question of Operational Taxonomic Unit arises again. It was noted that definition of OTU in for numerical taxonomy of urban form follows in principle *mixture problem* (see section 3.1.2.3), meaning that an OTU is necessarily an aggregation of fundamental elements. Moreover, it has to be area-based aggregation to avoid overlaps of elements brought by location-based techniques.

Assuming that the initial pool of fundamental elements is a *mixture* of separate “*populations*” which need to be identified, the population is a group of individuals of a single species. Even though the term species is abstract in urban morphology, its phenetic definition is very much applicable. Following @sneath1973, species is "*the smallest (most homogenous) cluster that can be recognised upon some given criterion as being distinct from other clusters* (p.365). That definition is conceptually the same as the definition of urban tissue in urban morphology as both are primarily based on the distinctiveness of each group identified either as species or as a tissue. However, a species is a taxon; therefore, its counterpart would be urban tissue type.

Therefore, within the framework of a mixture problem, we can consider urban tissue type an Operational Taxonomic Unit of numerical taxonomy of urban form.

### Analytical frameworks of urban form

Analytical frameworks of urban form are conceptual schemas linking fundamental elements together for the purpose of morphological analysis. The way we link fundamental elements and their aggregation’s matters and frameworks are often not specified in literature, just assumed.

The literature mostly offers frameworks which are hierarchical in nature, meaning that on a single level, each element can be part of a single aggregation. One of them is a framework used by @feliciotti2016 and @dibble2017 (figure ) which stacks elements together to form Sanctuary areas. A similar concept is used by @kropf2017 in his multi-level diagram of built form (figure ), which itself is based on older Canniggian theory [@caniggia1979].

Hierarchical framework combining individual plots into street edges, blocks and sanctuary areas. Reproduced from Feliciotti (2018).

Multi-level diagram of built form proposed by Kropf (2017, figure. 6.39)

A structurally different approach is proposed by @alexander1966. In his work *A City is not a Tree*, Alexander points out that city does not work in a tree-like hierarchy, which is reflected in both frameworks above, but as a semi-lattice of connections (figure ). However, none of the analytical frameworks tends to reflect that. Furthermore, Alexander’s own work on pattern language does not help in morphometric analysis.

Diagrams comparing the tree-like hierarchical structure (left) and overlapping semi-lattice (right). Reproduced from Alexander (1966).

Some frameworks focus on specific elements of urban form, but elements from different approaches are usually not linked together. A typical example is network analysis based on nodes and edges and its relation to hierarchical frameworks mentioned above. The value of, say ’degree of a node`, is not possible to reflect within the framework as the relation between a node and other elements is not captured.

While the literature shows that there is a broad spectrum of elements and their aggregations useful for morphometric analysis, no method links then into a singular overarching framework, which would, in turn, take a structure of semi-lattice instead of simple tree-like hierarchy. The key question here is how to build a framework which is not strictly hierarchical?

## Theory of Relational framework of urban form

To put things forward, this research proposes a relational framework of urban form for urban morphometrics.

### Relational analytical framework as an analytical tool

Relational analytical framework (RF) of urban form is based on two concepts - topology and inclusiveness. The framework acknowledges that there are identifiable relations between all elements of urban form and their aggregations. As such, it accommodates all analytical aggregations into a singular framework, linking all potential measurable characters to the smallest element. Furthermore, it employs topological relations in the way it generates location-based aggregations of fundamental elements.

Unlike frameworks above, relational analytical framework is analytical, not conceptual or structural. It does not try to propose a new theory of urban form; it has purely morphometric nature.

Within this research, relational analytical framework is operationalised based on morphological tessellation. That does not have to be the rule, but only one interpretation of the principle. The same could be done with plots and different aggregation frameworks.

#### Tessellation-based relational framework

Tessellation-based relation framework starts from two hypotheses. First, that morphological tessellation can be the smallest spatial unit in the morphometric analysis. Second, that tessellation-based contiguity aggregation is better than any other location-based aggregation framework. Both hypotheses will be further tested before the application of relational framework.

Assuming both hypotheses hold, the key principles are as follows.

1. Urban form is represented as building footprints, street networks and footprint-based morphological tessellation.
2. There is an identifiable relationship between buildings and street networks, buildings and street nodes and buildings and tessellation cells.
3. Morphometric characters are measured on scales defined by topological relations between elements.
   * Element itself
   * Element and its immediate neighbours
   * Element and its neighbours within n topological steps, either in a constrained or an unconstrained way.
4. Therefore, we can define subsets of relational framework as measurable entities of urban form based on fundamental elements and topological scales.
5. Subsets are overlapping, reusing each element within all relevant relations.

Since the relation between all elements is preserved throughout the process of their combination, we can always link values measured on one subset to another. For example, due to the fixed relation between building and street node, we can attach a node’s degree value to a building as an element. The constrained topological relation can identify traditional area-based aggregations like block (as a combination of all tessellation cells which topological relation does not cross a street). As such, they allow us to combine both area-based and location-based aggregations while minimising MAUP for each of them.

##### Subsets of elements

Subsets are a combination of topological scales and fundamental elements. Overlap of morphometric characters derived from subsets, where each subset is representing a different structural unit, gives an overall characteristic of each duality building - cell, which can be later used for further analysis.

We can divide subsets into three topological scales: Small (or Single), Medium and Large.

Note that topological distance is possible to define within each layer (relations between buildings, relations between cells, relations between edges or nodes), but not as a combination of layers. The relation between building, its cell, its segment and its node is fixed and seen as a singular feature. That is why morphometric characters like covered area ratio of the cell are classified as a Small scale character.

###### Small/Single (S)

Small scale captures fundamental elements themselves (topological distance is 0 - itself). In the case of building and tessellation cell, it captures the individual character of each cell. In the case of street segment and node, it captures value for segment or node, which is then applied to each cell attached to it.

We have four subsets within small scale:

* building
* tessellation cell
* street segment
* street node

Diagrams illustrating the subsets on the small/single scale.

###### Medium (M)

The medium-scale reflects topological distance 1. It captures individual character for each element derived from the relation to its adjacent elements.

* adjacent buildings
* neighbouring cells
* neighbouring segments
* linked nodes

Diagrams illustrating the subsets on the medium scale.

###### Large (L)

Large scale captures topological distance 2-n. In the case of cells, it captures individual character for each cell derived from the relation to cells within set topological distance. In the case of joined buildings and block, resulting measurable values are shared among all elements within such a structural unit. Block here is based on morphological tessellation and is defined as the contiguous portion of land comprised of cells which are normally bounded by streets or open space.

* joined buildings
* neighbouring cells of larger topological distance
* block
* neighbouring segments of larger topological distance
* linked nodes of larger topological distance

Diagrams illustrating the subsets on the large scale.

The resulting combination of all subsets is overlapping, following, in principle, Alexander’s schema more than hierarchical frameworks.

Diagrams illustrating the overlapping nature of the relational framework. The left diagram overlays all subsets on top of each other capturing the importance of each element for description of urban form around the indicated building. The darker the colour is, more times each element is used within various subsets. Diagram on the right shows all subsets aligned on top of each other describing the similar information while showing each subset directly.

## Method of tessellation testing

Other sections of this chapter test the viability of morphological tessellation as the smallest unit based on its ability to reflect the similar information as the plot and topological aggregation compared to other options. The relational framework itself can be tested only in an application, which is left for chapters 7 and 8.

### Tessellation as a unit

The first section tests whether tessellation holds as the smallest spatial unit, alongside the traditionally used plot, here represented by cadastral layer (i.e. following ownership-based definition of the plot). In the first part it presents a method of creation of morphological tessellation and in the second a method of assessment of resulting geometry in comparison to the plot.

#### Generation of Morphological Tessellation

Whenever observing a map or a satellite view of a city, the eye of the observer is caught by the existence of a fundamental relationship between buildings – their geometry and spatial configuration – and the plot pattern. This “*intuitive*” relationship is the reason why approaches based on Voronoi tessellation appear to *‘make sense*’ when applied to the urban form of cities: by partitioning the space into cells, they capture the way buildings relate to each other in space and, more precisely, give a spatial meaning to the “*morphological influence*” that each building exerts on its immediate spatial context [@usui2017]. It, in turn, implicitly captures how spatial configuration affects visibility, light penetration, ventilation, movement, etc. around each and every building [@hamaina2012a].

The main advantage of methods based on Voronoi tessellation is the capacity to derive objective spatial partitions that are applicable to every type of urban tissue in a way independent from the researcher’s subjective interpretation. In addition, most of these methods [@hamaina2012a; @hamaina2013] require minimum data input, as they fundamentally rely on the polygon that describes the footprint of a building. Similarly, the proposed morphological tessellation method only requires a polygon layer representing building footprints (figure a). From this, morphological tessellation moves forward in five steps:

1. *Inward offset from building footprint* (figure b). The offset is necessary to avoid overlaps between boundaries of adjacent buildings and generate a gap between adjacent geometries which will later define the boundaries of the cell.
2. *Discretisation of polygons’ boundaries into points* (figure c). As Voronoi tessellation can efficiently be generated only from point features, the polygonal shape of the building footprint needs to be approximated as series of points to be placed at regular intervals along its boundary, where generated points retain the ID of the building they belong to.
3. *Generation of Voronoi cells* (figure d). Voronoi cells are generated around each of the points representing the building footprint. Again, the original ID of the building is preserved in the resulting VC.
4. *Dissolution of Voronoi cells* (figure e). All Voronoi cells sharing the same building ID – and hence generated from the same building – are dissolved in unitary geometries. This step provides a preliminary boundary of the morphological cells.
5. *Clip of preliminary tessellation* (figure f). As a geometrical construct, Voronoi cells tends to infinity as the boundaries of each cell are only defined by proximity with adjacent *“seeds”*. However, when applied to the analysis of urban form, for obvious reasons, no cell can tend to infinity. To avoid this, it is necessary to limit the maximum spatial extent of the tessellation by setting defined study area boundaries.

The proposed morphological tessellation method. Grey polygons represent building footprints, while red lines show the edges of tessellation at each step. 3a) Building shapes within the boundary of the study area (blue); 3b) inward offset from building footprint polygon; 3c) discretisation of boundaries of polygons into points; 3d) generation of VCs around points: at this stage, the edges of cells (red) tend to infinity; 3e) dissolution of Voronoi cells based on original building ID; 3f) clip of preliminary tessellation by study area.

Three of the five steps listed above, namely inward offset distance (step 1), discretisation interval (step 2) and clipping method (step 5), require setting parameters that can have a significant effect on the resulting tessellation. As such, these need to be evaluated in greater detail. More specifically, in the case of inward offset distance (step 1), the selection of too large values may cause the collapse of narrow parts of building shapes and loss of detail, while too small ones may generate unwanted “*saw-like*” geometries between adjacent buildings. Similarly, a large discretisation interval (step 2) may produce the same “*saw-like*” geometry issue, whilst the opposite would increase exponentially computational demand (figure ). Additionally, since the two parameters are interlinked, their individual effect on the shape of each cell is not independent: as such, their combined effect needs to be balanced to generate geometries with insignificant shape deviation and minimum computational burden. Finally, the adoption of a clipping method for the tessellation (step 5) also requires considerations in order to appropriately limit the focus of the analysis to the urbanised footprint and exclude large open un-built spaces while limiting potential MAUP effects [@openshaw1984]. Due to the importance of correctly setting these parameters, section 6.3.1.1. will discuss the adopted method for the determination of inward offset distance (step 1), discretisation interval (step 2) and method.

Illustration of the effect of an improper combination of inward offset distance and discretisation interval causing “saw-like” geometry on the boundary between adjacent buildings (b) compared to ideal combination (a).

The conceptual sequence described in this section was translated into a Python code, building its key parts on the capability of SciPy [@jones2001], Shapely [@gillies2007] and GeoPandas [@kelsey\_jordahl\_2020\_3946761]. Computation was run on Ubuntu Bionic 18.04 running at Amazon Web Services EC2. The resulting Python script is released as part of momepy.

#### Morphological Tessellation and plots: data and comparison method

##### The dataset

Even though the rest of the thesis works primarily with Prague and Amsterdam as a case study, the following section focuses on the administrative area of Zurich, Switzerland (figure ). It was chosen for its historically characterised and heterogeneous urban fabric as well as for the availability of the “*Amtliche Vermessung*” dataset, a freely-accessible resource containing high-quality information on cadastral plots and building footprints. Before generating the morphological tessellation, data was cleaned as follows:

1. From the *cadastral layer*, which covers the 100% of the study area, all features not containing buildings (e.g. streets or large open spaces) were removed, as they do not represent built-up form;
2. From the *building layer*, features smaller than 30 m2 were filtered out, as such smaller objects are likely ancillary structures rather than actual buildings.

The selected study area, defined by the administrative boundary of the Zurich Kanton (left); the Langstrasse area in Zurich (right) was selected for testing the tessellation algorithm parameters: the red-line boundaries follow the street centerlines.

##### Definition of morphological tessellation

parameters: inward offset distance, discretisation interval and clipping method

To determine the optimal setting for inward offset distance (step 1) and discretisation interval (step 2), a test was run on a portion of the Langstrasse area in Zurich (Figure 5), a heterogeneous fabric predominantly characterised by adjacent buildings (significantly more prone to error than isolated buildings) limited by the street network. The test considered several combinations of inward offset (from 0.1 to 1 meter) and discretisation interval (from 0.05 to 5 meters) and evaluated them against the most precise setting (0.1 / 0.05), which provides the highest-resolution tessellation with minimal effect on the building shape. The test then assessed deviation of cell perimeter and area values for each combination, as well as its computational demand: the latter is a function of the number of discretisation points, as these directly impact on memory and processing demand. The result of this test is presented in section 6.4.1.1. Based on it, the optimal combination of the two parameters was adopted to generate the morphological tessellation in later stages.

Finally, in order to clip the tessellation (step 5), the test adopted a definition of urban footprint aligned with Angel et al. [-@angel2007; -@angel2018], and limited the study area extent by setting a 100m buffer from the built-up area. However, to test the robustness and stability of the buffer and avoid arbitrary selection, it also tested 14 other buffers, ranging from 10 to 300 metres. The stability of the 15 buffer distances is discussed alongside the comparative analysis in section 6.4.1.2.

##### The informational value of morphological tessellation

vs plots: the 12 morphometric characters

To test the informational value of the morphological tessellation compared to plots, 12 morphometric characters (table ) grouped into the six morphometric categories (dimension, shape, spatial distribution, intensity, connectivity and diversity) proposed in chapter 4, are selected and measured on both the cadastral layer and morphological tessellation layer, at the 15 buffer distances. With the exclusion of Reach Centrality, which is measured using the *Urban Network Analysis (UNA) Toolbox* [@sevtsuk2012], all characters are computed using Python scripts released as part of the momepy package.

Once all morphometric characters are calculated for cadastral plots and the 15 morphological tessellation layers (at each buffer distance), the similarity of the resulting values for the two datasets is evaluated using three methods: 1) Spearman’s rank correlation; 2) Normalised root squared mean deviation (NRSMD) and 3) Accuracy of significant patterns defined by local Moran’s I indicator of spatial autocorrelation (LISA) [@anselin2010]. Spearman’s rank correlation is “*a measure of the correlation between ranks, calculated by using the ranks in place of the actual observations in the formula for the correlation coefficient r*” [@kokoska2000, p.372] (see Equation 1) and was used due to non-normality of distribution of measured values. It ranges from -1 (negative correlation) to 1 (positive correlation), with values > 0.5 or < –0.5 indicate moderately significant positive or negative correlation [@hinkle2003].

1. $\r\_{s}=1-\frac{6 \sum d^{2}}{n\left(n^{2}-1\right)}$,

where is the difference between the rank of observed and expected value and is the number of observations [@kokoska2000].

NRSMD is a frequently used measure of “*an estimate of the standard deviation of residuals from the model*” [@alexander2015, p.5] normalised by the range (see Equation 2), and it is used to measure the difference between the expected and observed values, normalised by the range. As a ratio of deviation, it ranges from 0 to 1, where 0 means no deviation and 1 means deviation equal to the range of values. As the range is sensitive to outliers, NRSMD might not be relevant for characters of *Dimension* category.

1. ,

where where are observed and expected values.

Accuracy is “*closeness of computations or estimates to the exact or true values that the statistics were intended to measure*” (OECD, 2006). It is here used to measure the similarity of significant spatial clusters identified from the cadastral layer and those identified from each version of the tessellation (see Equation 3). Since studies in Urban Morphometrics are more interested in uncovering *recurrent patterns* in urban form rather than *actual values* [@feliciotti2018], this method is probably the most relevant of the three. In fact, it measures whether corresponding features from both datasets (cadastral plots and morphological cells in this case) significantly fall within the same cluster (i.e. p <= 0.05), with values ranging from 0 (no match) to 1 (perfect match).

1. ,

where is the number of the elements belonging to the same significant spatial cluster in both and is the number of the elements belonging to any significant cluster. The adjacency matrix used for LISA represents 200 metres Euclidean distance from each building.

It must be noted that, for the statistical comparison of selected morphometric characters across the morphological tessellation layers and the cadastral layer, these must correspond perfectly. However, whilst there is a 1:1 match between morphological cells and *buildings*, the same does not apply to morphological cells and plots, as the latter may contain one building (single-building plots) or more than one (multi-building plots). To resolve this issue, the building layer is used as a proxy between tessellation and cadastre and, therefore, all morphometric characters computed on both morphological cells and plots are associated to the building layer (i.e. each building is linked to the value of its morphological cell and of the plot it sits on). However, to better understand the impact of “*single-building*” and “*multi-building*” plots (79% and 21% of all plots respectively), the three methods described above are applied to the whole dataset and, separately, for *single-building* and *multi-building* plots. In particular, we expected that *multi-building* plots, although important for their effect on the overall analysis, would hold limited comparative value for most of the assessed morphometric characters (perhaps with the only exclusion of covered area ratio and Gini index of CAR, which capture compatible concepts).

### Topological contiguity of tessellation as an aggregation framework

The second test of tessellation within this chapter focuses on its ability to derive topological location-based aggregations. The definition of aggregated analytical units via the topology of morphological tessellation can overcome issues of the three methods described above and provide a more consistent way to understand the relationship between adjacent elements of urban form (in the case of buildings or morphological cells). This section is testing this hypothesis in the case of Prague, Czechia.

#### Comparing aggregation methods

The methodology of this research follows a twofold approach, analysing both small scale case studies and urban scale statistical data. Small scale case studies examine the difference between three methods extracted from literature (Euclidean distance, metric reach, K-nearest neighbour) and unconstrained topology of morphological tessellation in different types of urban form; large scale statistical analysis examines the parameters of these methods of aggregation across the whole of Prague.

The method compares how each of the tested methods aggregate tessellation cells (being smallest spatial unit) within two scales: one achieved by nine topological steps, the equivalent of approximately 400 metres used in the morphological analysis to represent a walking distance of 5 minutes; and one achieved by 4 topological steps, representing roughly 200 metres. The number of neighbours for KNN is then derived from the mean number of neighbours captured by each of the topological distance and metric reach, to keep the dimension comparable (table ).

Default topological distances and their equivalents. Values are derived from the summative analysis of topology-based aggregations defined around each morphological cell on Prague.

Topology of MT

Euclidean

Metric reach

KNN

4 steps

200 metres

200 metres

70 neighbours

9 steps

400 metres

400 metres

320 neighbours

: Default topological distances and their equivalents. Values are derived from the summative analysis of topology-based aggregations defined around each morphological cell on Prague.

The aim of the small-scale analysis is to understand how each of the four methods identifies and represent the same information across 5 types of urban form - medieval organically grown, 19th century compact perimeter blocks, 20th century mixed single and multi-family villas, 20th century modernist housing, and 20–21th century industrial estates. It should allow ascertaining how consistently each method distinguishes variations in form and morphological behaviours.

At the urban scale is conducted a statistical analysis to compare the distribution of values derived from the whole of Prague. The statistical distribution of data across the whole urban area describes the spread and variance of values, which can be used to assess the ability of each method to capture the intended information across types of urban tissue. To understand the different performance of each method, the method compares distributions of two descriptive variables as a proxy for the performance assessment – number of neighbours and covered area.

The first variable is the number of neighbours captured. Neighbours represented by buildings and related tessellation cells capture most of the morphological information. For that reason, it is desirable to use the method which will identify the somewhat similar number of neighbours no matter the urban tissue to keep the similar essence and amount of information to maximise comparability of values. It means that the distribution of such values should have a relatively small standard deviation and be close to symmetrical distribution to have a similarly positive and negative deviation from the mean.

The second variable should represent the concept of the geographical extent of aggregation, as it bears the information of the scale of each type of urban tissue and therefore could describe the ability of each method to adapt to the scale. Amongst the possible measurable variables are mean distance to neighbours, maximum distance to neighbours and area covered by aggregation. Because they all represent the scale and extent of aggregation of elements (buildings, tessellation cells), we use the only area covered to represent them all as it is the most straightforward one and easy to understand. The statistical distribution of the covered area should represent the adaptability of the aggregation method. Hence, the ideal outcome should have a high standard deviation and high range of values, meaning that many different options (levels of granularity of urban form) are all captured.

## Results

The following section presents results of the analysis, first assessing the tessellation as a unit in the case of Zurich and then analysing its aggregation ability in the case of Prague.

### Tessellation as a unit

Test of tessellation as a unit focused on two aspects - determination of optimal parameters of the algorithm generating tessellation from building footprints and comparison of morphometric values between the cadastral layer and tessellation.

#### Determination of optimal parameters of the MT algorithm

The test performed on the selected inward offset ranges (from 0.1 to 1 meter) and discretisation intervals (from 0.05 to 5 meters) allowed to assess computational demand (i.e. a number of discretisation points) and deviation of cell perimeter and area for each combination. In terms of computational demand, as shown in figure , it appears that the discretisation segment length has an exponential effect on the number of generated points. For values below the mean (tail of the distribution), computational demand remains relatively stable, whilst for higher values (head of the distribution) it grows sharply, more than doubling at each step. Discretisation intervals ≥ 0.5m are therefore preferred as more computationally effective.

Relation of discretisation segment length and number of points generated. The red line illustrates the mean value above which the number of points more than doubles at each step

The effect of negative buffer and discretisation interval on the deviation of the morphological cell’s area compared to the high-resolution tessellation is insignificant for all tested combinations (0.00 and 0.01%), showing that, no matter the parameters, results are stable. In turn, the same effect on the morphological cell’s perimeter is more pronounced (figure ) due to the aforementioned phenomenon of “*saw-like*” geometries (see the section 6.3.1.1) with per cent deviation ranging from 0.05% to 7.4%. Focusing on the 0.5 metres discretisation interval, providing the balance between the morphological cell shape detail and computational demands, deviation values range from 0.47% to 3.1% (figure ). It suggests that the combination of 0.5m metres discretisation interval and 0.4m inward offset distance provides the optimal balance in terms of the effectiveness of computation and minimisation of error. These values are hence adopted as parameters in the computation of the morphological tessellation in the next stages.

The mean deviation in percents of perimeter of each cell for each combination of inward offset distance (vertical axis) and discretisation interval (horizontal axis).

Relation of inward offset distance and error margin, showing that for 0.4 meter, the error margin reaches its minimum.

#### Comparison between the cadastral layer and morphological

tessellation

Having determined the optimal combination of the tessellation parameters (inward offset = 0.4m and discretisation intervals = 0.5m), the morphological tessellation for Zurich is computed using momepy. From a first visual inspection of the generated layer, it is already possible to appreciate how the morphological tessellation is able to capture variations in size nicely, grain and compactness of buildings (figure ), not dissimilar what observed in a typical cadastral layer. The method subsequently calculates the 12 morphological characters in table for cadastral plots and tessellation cells. In the next section, their correlation at each buffer of tessellation is studied.

Morphological tessellation cells as generated across four different areas of Zurich; 4a) organic tissue of Niederdorf; 4b) compact tissue of Langstrasse; 4c) detached villas of Hottingen; 4d) mixed post-war development of Friesenberg.

##### Spearman’s rank correlation

Using Spearman’s rank correlation, results show that correlation of measured characters ranges between 0.25 (fractal dimension) to 0.89 (reach), with differences between morphometric categories and between *single-* or *multi-building* plots. Characters in the *Shape* category exhibit the worst performance, with insignificant correlation for the whole sample (~0.27) and multi-building plots (~0.09) and low significance for single-building plots (~0.42). This result was expected, due to the intrinsically different geometry of the two spatial units (morphological cells and plots) and to the existence of multi-building plots.

*Dimension* characters inherently differ between *multi-* and *single-*building plots, showing only low significance for the former (~0.35, ~0.4) and high significance for the latter (~0.83, ~0.7). Remaining characters show moderate or high significance for all samples, with higher values for single-building plots (figure and Table 2). Results for all buffers are relatively consistent, with fluctuations observed only at smaller distances (< 50m), indicating the stability of the selected value of 100m.

Spearman’s rho rank correlation between cadastral values and each of the selected buffers of tessellation based on the whole dataset (figures for single and multi-building plots are found in Appendix 6.1).

##### Normalised RMSD

Overall, the RMSD test indicates a high level of similarity between datasets (figure and Table 3), excluding *Dimension* characters which, as mentioned in Section 3.2 are heavily skewed by large outliers, hence not comparable with the rest of the data. Apart from *Orientation*, which is the worst-performing character in the set (~0.22 for the whole dataset, ~0.26 for multi-building plots and ~0.18 for single-building plots), all other characters score RMSD values lower than 0.2 (~0.15 for single-building plots and ~0.18 for multi-building plots). It suggests that, even though the spatial coverage of the morphological tessellation is different from plots, this difference is, in terms of information, only minor. Even the poorer performance of *Orientation* depends more on the way this is measured than on dissimilarity between datasets: unlike other metrics, *Orientation* is calculated as a deviation of the orientation of the longest axis of minimum bounding rectangle (sometimes called oriented envelope) from cardinal directions in degrees and, as such, it ranges from 0 to 45º. Hence, a deviation of 0.2 corresponds to a difference only of 9º. It is worth noting that for smaller buffers (15 to 40m) results show high instability, where some characters exhibit the highest correlation values and others the lowest: this confirms that smaller buffers are unsuitable as parameters to limit the tessellation. In turn, the 100-metre buffer is confirmed as robust and stable across all characters.

NRMSD of cadastral values and each of the selected buffers of tessellation based on the whole dataset (figures for single and multi-building plots are found in Appendix 6.2).

##### Recognition of significant patterns using LISA

Example of LISA patterns of Frequency measured on the MT shows Moran Local Scatterplot LISA clusters and actual distribution of values.

The analysis of patterns with LISA (figure ) captures differences across measured characters; however, given the dissimilarity of the datasets due to multi-building plots, the accuracy scores are not expected to reach values close to 1. The highest pattern similarity is recognised for the *Frequency* character (~0.78, corresponding to an almost 80% match) (figure ), followed by *Orientation* and *Diversity* characters (Gini Index of Area and Gini Index of CAR) and CAR (0.74 – 0.66), while *Dimension* characters are around ~0.5 depending on the sample considered (single- or multi-building plots). *Shape* characters are consistently the ones providing lowest accuracy, apart from *Reach Centrality*, due to the single-building vs. multi-building deviation in the datasets. Overall, the difference between samples is relatively consistent, with single-building plots reaching values between 0.1 and 0.2 higher than multi-building plots. While none of the values indicates an equality of both datasets, some are close enough to be considered as proxies of each other. The effect of buffer distance confirms already observed pattern and the stability of the 100m buffer.

LISA accuracy of cadastral values and each of the selected buffers of tessellation based on the whole dataset (figures for single and multi-building plots are found in Appendix 6.3).

### Tessellation contiguity as an aggregation method

At the small scale, the method studied how five urban types in Prague are represented by the four methods of aggregation. The Old Town, a tissue of a medieval origin which has grown organically, shows few differences between the four methods, with slightly larger footprints of aggregations defined by Euclidean distance and a morphological tessellation topology at both 200 (figure a) and 400 (figure a) metres distances. Numerically, the difference is clear, but for pattern-detection this difference is not substantial, suggesting that all methods are relatively equal in this tissue for both 200 and 400 metres (and equivalents). Differences might be explained by the high granularity of the tissue, with many elements on a relatively small area (the reason for KNN being the smallest) and complex configuration amongst them (buildings have many neighbours, expressed by more extensive topological-based aggregation).

Comparison of boundaries of aggregations defined by each of the tested method for 4 topological steps and equivalents (200 metres, 70 neighbours). a) Old Town, b) Vinohrady, c) Hanspaulka, d) Jižní Město, e) Malešice

In the second case, the urban tissue of 19th century of compact perimeter blocks in the Vinohrady neighbourhood results almost match the previous case. Due to the high granularity of this urban tissue, purely Euclidean distance-based area is the largest, while K-nearest neighbour the smallest (figure b, figure b).

Overall, in these two historic tissues, the difference is not significant to conclude that one method is better than the other - simple visual comparison shows that boundaries almost overlaps for both tested distances.

The first crucial differences are noticeable in Hanspaulka, an area of 20th century mixed single and multi-family villas, where the street network is becoming less dense and less connected than in the city centre, leading to the difference between the area captured by metric reach (smaller) and the other three methods, which almost overlap without any substantial distinction (figure c, figure c). This indicates that the street network plays a crucial role in the applicability of reach-based methods.

Comparison of boundaries of aggregations defined by each of the tested method for 9 topological steps and equivalents (400 metres, 320 neighbours). a) Old Town, b) Vinohrady, c) Hanspaulka, d) Jižní Město, e) Malešice.

The twentieth-century modernist housing of Jižní Město is the first example of a post-WW2 urban tissue. The planning ideology behind it comes with the radical change of scale and a distinctive approach to streets and their connectivity. This change is reflected in how each of our tested methods captures the space: while the area defined by Euclidean distance remains mostly the same as in pre-WW2 tissues, the area captured by metric reach shrinks due to the convoluted street network. In theory, both topological and KNN definitions of aggregation should be able to capture the difference in scale and up to a certain level they do. However, KNN, even though being larger than metric-based methods, lacks the ability to deal with large pavilion-like buildings with many direct neighbours, unlike the topological definition which correctly reacts to the abrupt change of scale of the granularity of urban tissue and captures the relationship between high-rise buildings and their low-rise pavilion counterparts by acknowledging that they are neighbouring (figure d, figure d).

Whilst industrial type tissues are generally not the concern of urban morphology, as classified as specialist and treated differently than more ordinary fabric, they are nonetheless large, therefore important parts of our cities and as such deserve to be studied using the same approach as the more conventional ones. Their scale is radically different. Buildings are of the size of the traditional block or larger, the plot structure is mostly unorganised, and the street network is utilitarian only, following different principles than in residential or mixed-use parts of the city. These differences are captured through the application of our four methods. The network-based method is unreliable on this tissue, capturing the only a minor area around the building due to the major drop in a granularity and connectivity. The Euclidean distance of 200 or 400 metres, which seems to capture enough information in more granular urban tissues lacks the same capacity in this case. K-nearest neighbour analysis struggles to capture the peculiarities of this particular urban tissue, which is characterised by a large amount of additional built-up structures to main buildings, leading to the identification of smaller area that makes a comparison with the other cases confused. The topological definition achieved by the morphological tessellation seems to tackle all issues of the other methods, whilst capturing a similar amount of information as it did in previous cases (figure e, figure e).

Overall, the differences between methods in defining aggregation are heavily dependent on the type of urban tissue analysed. More traditional (from a European perspective) urban tissues like medieval (Old Town) or perimeter blocks (Vinohrady) indicate that in these contexts the choice of the method is purely the matter of opinion and that the resulting value offered by the four methods is mostly similar. However, once we start focusing on post-WW2 development, we often observe a change of scale of urban patterns, which makes distance-based methods (Euclidean, metric reach) unable to react to such change. The information captured is consequently different in pre-and post-WW2 urban tissues, complicating the further comparability, whilst we seek similar and consistent data. If the urban patterns change their scale, the method of capturing such an extent needs to be able to adapt to it. Our results of the small-scale analysis indicate that topology of morphological tessellation is the method able to fulfil this condition adequately. Whilst small scale analysis illustrated the capacity of the four selected methods to provide stable information, it is only at an urban scale, through statistical analysis, that we can show a full overview of how the four methods perform.

As mentioned, neighbouring elements are bearing the primary information about urban patterns. For this reason, researchers aim to use methods capturing an equal number of neighbours across contexts. Such a method might be K-nearest neighbour, but due to the variety of urban configurations, a method needs a certain level of adaptability (which KNN with a fixed number cannot provide). As figure (a, b) shows, the statistical distribution of the number of neighbours captured is the most stable for the topology of morphological tessellation, being almost perfect Gaussian distribution (the deviation in the number of neighbours is the same in both directions from the mean), with the smallest standard deviation (). The metric reach method to provides right-skewed distribution and Euclidean distance high deviation, which are both undesirable features in terms of stability of information.

Then, the comparison of distributions of covered area aims to test the adaptability of each method. As mentioned, the changing scale of urban patterns means that the same level of information is spread to larger areas. Therefore, an ideal method should show high flexibility (the distribution of values should have large range and high standard deviation) in the area captured to fit all patterns possible. The results as shown in figure (c, d) indicate that topology of morphological tessellation offers by a large margin the highest standard deviation out of tested methods, indicating that the change of the scale is captured successfully. Metric methods (Euclidean distance, metric reach) are the least flexible in this sense, while K-nearest neighbour might offer desired value alongside with morphological tessellation topology.

Statistical distributions of number of neighbours a) 4 steps and equivalents, b) 9 steps; and total covered area c) 4 steps and equivalents, d) 9 steps.

Even though there are differences between smaller and larger distances (4 steps / 9 steps and its equivalents), topologically defined aggregation seems to reflect desired outcomes (i.e. stability in a number of elements captured and flexibility in metric values) in both statistical comparisons better than other tested methods. This finding is in line with the one we drew from small-scale case studies, indicating that topology of morphological tessellation is a valuable approach to be employed in morphological analysis.

## What is the value of tessellation?

Results of both experiments show that there is a high potential in using morphological tessellation as a spatial unit, both as a unit of analysis and as a core of the aggregation technique.

### Tessellation in relation to plot

The results suggest that the proposed method contributes to resolving some of the limitations associated with using the plot as a unit of morphometric analysis. However, the picture resulting from testing the similarity between cadastral plots and morphological cells is rather complex. Notably, the significance of the similarity between plots and morphological cells varies considerably depending on the morphometric character selected: this is generally high for all Intensity characters (*Frequency, CAR*, whilst Shape characters (*Rectangularity, Circular Compactness, Shape Index*, and *Fractal Dimension*) report a comparatively lower performance and a higher deviation. It means that if, for several of the morphometric characters assessed, morphological tessellation is able to retain plot-level information which is comparable to that provided by the cadastral layer, for other characters morphological cells are less efficient proxies of plots and capture comparatively different information.

It is also evident that the similarity of datasets is higher across all measured characters for single-building plots compared to multi-building plots. To consider that the former ones are predominantly found in pre-industrial urban tissues. At the same time, the latter is more typical of modern and contemporary development, and it is suggested that the morphological cell might capture similar information as the plot in the context of traditional fabrics better than in modernist and contemporary ones.

Overall, there appears to be a scope for the morphological cell to be utilised as the basic unit of morphometric analysis, given its ability to capture meaningful patterns of urban form at the plot scale, the degree of reliability and universality of the underlying method and the wide accessibility of the data required to generate it. Indeed, while the recognition of plots can be very troublesome and resource-intensive, morphological tessellation is consistent throughout, since it is only based on building footprint information which is equally present in all kinds of urban areas. Moreover, by using morphological tessellation instead of traditional methods relying on buildings, street networks and plots, data dependency is reduced by a third as the tessellation is generated from the building layer alone.

On the other hand, the morphological tessellation cannot fully substitute the plot. The similarity of information indicates only partial overlap, and there are aspects of form which cannot be derived from morphological tessellation conceptually, like the relationship between public and private or ownership-based analysis. That said, the morphological tessellation has the potential to work as an analytical spatial unit within the morphometric assessment, acknowledging that the information obtained from it is not equal to that derived from plots (no matter the definition).

### Location-based aggregation using tessellation

Existing location-based methods of aggregation of elements into larger analytical units all face some issues limiting their applicability and reliability. The alternative method presented is based on the topology of space as captured by the morphological tessellation. Such a method of partitioning space reflects the influence of each building on the space around it to overcome existing challenges and provide a context-sensitive method. Initial results of the twofold analysis of the topological ability of the morphological tessellation indicate that the type of urban tissue influences the outcome of morphological analysis and that in the case of pre-WW2 traditional European-like urban tissues, all currently available methods of definition of aggregation are relevant and almost interchangeable. However, this is not the case with post-WW2 urban developments, as in them there has been a significant change in the scale of form’s granularity. In these cases, urban morphology needs to employ methods which are sensitive to the scale and configuration of urban form and at the same time can detect its granularity. The morphological tessellation and the topology derived from the analysis of its structure seem to be the most successful, sensitive method, suitable for general analysis. All of the methods that have been tested partially solve one of the key issues identified in spatial analysis (MAUP), as data are aggregated independently for each element, and there are no preselected boundaries in play.

No matter the results of the presented analysis, methods extracted from the literature have their role in morphological analysis. However, Euclidean definition and metric reach should be used in specific situations only, due to their limitations, as mentioned above. It is either in stable environments without abrupt changes of granularity or in a definition of larger-scale aggregations, where multiple urban tissues are included. In that case, the main benefits of the morphological tessellation — following the spatial configuration of urban patterns — is not so crucial and from certain scale does not even provide added value.

For the analysis done on the small scale (scale of urban tissue and smaller), Euclidean definition and metric reach do not provide stable information, unlike KNN (which always captures the same number of elements of similar informational value, but not the same relationship) and the topology of morphological tessellation. Moreover, aggregation defined via the topology of morphological tessellation may be used even on the smallest scales of one or two steps as it will always capture intended comparable information based on the relationship of elements.

## Summary

The method of morphometric analysis presented in this thesis is designed with applicability in mind. To ensure that the paucity of viable input data does not limit it, the analytical framework is based on fundamental morphometric elements only. Those reflect fundamental elements known in urban morphology - building, street and, to a degree, the plot. Plots are commonly seen as the ideal spatial division for morphological analysis, but they also have their drawbacks, causing the limited applicability of plot-based methods and, more importantly, the reduced reliability of results obtained by employing them. This chapter tries to address some of the issues characterising the definition of the plot and plot boundaries, the availability and accessibility of plot data and the labour intensiveness of manually extracting reliable plot-level information, aspects that limit the potential of urban morphometrics. The need to objectively define a unit of analysis able to capture the smallest and arguably most fundamental level of spatial subdivision, and to develop a reliable and replicable method to generate and measure it, is the rationale behind the *morphological cell* unit and the *morphological tessellation* method.

The universal and algorithmic nature of the proposed morphological tessellation has the potential to scale up morphometric analysis with minimum effort to the large scale, while significantly reducing the interpretative input of the analysts along the process. This latter property of morphological tessellation appears to be particularly relevant to making large scale morphometric analysis viable and take full advantage of big data in the GIS area. The robustness of the proposed method and the validity of the proposed spatial unit of analysis is verified through the assessment of 12 representative morphometric characters and the application of three different quantitative comparative methods, Spearman’s correlation, NRMSD and accuracy of LISA, aimed at evaluating the similarity of information between morphological cells and cadastral plots.

The morphological tessellation, as tested and presented in this chapter, offers a different approach to spatial division whilst still capturing a level of quality of information on urban form that is, to a degree, similar to that conveyed by the plot. Findings presented in this chapter indicate that there is a partial overlap between the information derived from cadastral plots and the one derived from morphological tessellation. The degree of this overlap depends on the category of morphometric characters and the type of urban context, but for certain types of morphologic analysis, it is large enough to consider the information comparable to the one derived from plots. At the same time, it is important to keep in mind that morphological cells cannot fully *replace* plots in the understanding and analysis of urban form patterns.

However, the morphological tessellation is a step towards achieving consistency in urban morphology in both definitions of the smallest spatial unit and analytical aggregation. The advantage of morphological tessellation is that it limits the data dependency as it is based on building footprints only and allows the elimination of subjectivity in the partitioning of space. Most importantly, it is context-sensitive, allowing the researcher to use the same method across different types of urban tissues whilst still get comparable information, much needed for reliable results of any statistical analysis.

On top of an application of tessellation itself, this chapter proposes the tessellation-based relational framework for morphometric analysis of urban form based on the idea of overlapping elements and their aggregations. The resulting description is then based on the semi-lattice of relationships between individual subsets of measurable features. Since tessellation seemingly holds as a spatial unit, the proposed tessellation-based relational framework will be used and tested in the next chapters as a basis of morphometric analysis.

Finally, the overview of morphological aggregations and their linkage to a mixture problem of identification of OTU resulted in a working hypothesis of urban tissue types as OTU for numerical taxonomy of urban form. To which degree can we operationalise this idea and how to use morphometrics to delineate tissue types prior taxonomy are questions left for the next chapter.